

PROCESS FOR THE PURIFICATION AND CRYSTALLIZATION OF PROTEASOME

Description

The invention concerns a process for isolating a purified eukaryotic crystallizable proteasome preparation and the proteasome preparation obtainable by the process. In addition the invention concerns a purified eukaryotic proteasome preparation in a crystallized form. With the aid of the crystal data from this proteasome preparation it is possible to identify and obtain new proteasome inhibitors, especially with the aid of computer-aided modelling programs.

Proteasome is the central enzyme in protein degradation in the cytosol as well as in the cell nucleus. It is involved in many biological processes including the removal of abnormal, misfolded or falsely assembled proteins, the reaction to stress (by processing or degradation of transcription regulators), cell cycle control (by degradation of cyclines), cell differentiation and metabolic adaptation (by destruction of transcription factors or metabolic enzymes) and the cellular immune reaction (by generation of antigenic peptides that are presented by MHC class I molecules). These cellular functions which are based on a ubiquitin and ATP-dependent degradation of proteins require the 26S proteasome whose nucleus and proteolytic chamber is formed by the 20S proteasome.

The 20S proteasome from *Archaeabacterium Thermoplasma acidophilum* was analysed by X-ray structure crystallography at a resolution of 0.34 nm. It has a

cylindrical shape with a length of 14.8 nm and a maximum and minimum diameter of 11.3 nm and 7.5 nm respectively. It is composed of 28 subunits which are arranged in a particle as 4 homoheptameric rings $\alpha_7\beta_7\beta_7\alpha_7$ with D7 symmetry (Löwe et al., (1995), *Science* 268, 533-539). In the *T.acidophilum* proteasome the N-terminal threonine residue of the β subunits is the binding site of inhibitory peptide aldehydes and is essential for the hydrolytic activity. Stock et al. ((1996), *Current Opinion in Biotechnology*, 7: 376-385) also describe the structure and function of *T.acidophilum* proteasomes. The *T. acidophilum* data cannot be used for eukaryotic proteasomes since the homology of the proteasomes between these species is too low.

Eukaryotic proteasomes are considerably more complex than the archaebacterial proteasome. Thus the 20S proteasome from *Saccharomyces cerevisiae* is composed of a total of seven different α type and seven different β type subunits which have already been cloned and sequenced cf. e.g. Heinemeyer et al. (1994), Biochemistry 33, 12229-12237).

The eukaryotic 20S proteasomes e.g. from yeast and from mammals are very closely related with regard to the amino acid sequences of subunits and their coarse structure recognizable by electron microscopy. The α type and β type subunits of the mammalian 20S proteasome form an ordered and well-defined structure (Kopp et al. (1995), J. Mol. Biol. 248, 264-272). In mammalian cells three additional non-essential subunits of the 20S proteasome which are named LMP2, LMP7 and MECL1 can replace constitutive components after induction with the cytokine interferon γ . Their expression or specific deletion changes the peptidase specificity of the

proteasome and the expression rate of MHC class I molecules on the cell surface.

Hilt, Heinemeyer and Wolf ((1993), Enzyme Protein 47: 189-201) describe the structure of 20S and 26S yeast proteasomes and the proteolytic activity of β type subunits. In addition the different functions of 20S and 26S proteasomes are discussed with regard to the metabolism and differentiation of a cell. Crystallographic data of eukaryotic proteasomes are not described.

Starting materials that have previously been used to purify proteasomes such as tissue and cells of mammals such as mouse, rat, human or bovine, other animals, plants and yeast are listed in the publication by Rivett et al. (1994), Methods Enzymol. 244, 331-350) and the citations included therein.

There are also numerous documents in the patent literature which relate to proteasomes. Thus for example the production of eukaryotic proteasomes is described in EP-A-03 45 750, JP-A-05 292 964 and JP-A-06 022 759. However, the proteasome preparations that are disclosed do not have sufficient purity to enable crystallization.

Morimoto et al. ((1995), J. Biochem. 117, 471-474), Hwang et al. ((1994), Mol. Cells, vol. 4, 273-275) and Perkins et al. ((1994), Journal of Structural Biology 113, 124-134) describe the crystallization of eukaryotic proteasomes. Resolutions of only 0.44 nm, > ca. 5.0 nm and 1.5 nm respectively are achieved due to the low purity of the proteasome preparations so that a structural determination or molecular modelling is not possible with these proteasome preparations.

092041266-13007200

Nucleotide and amino acid sequences of proteasome subunits are described for example in the Japanese applications JP-A-04 077 497, JP-A-04 077 498, JP-A-04 117 283, JP-A-05 317 059, JP-A-07 255 476, JP-A-08 116 972, JP-A-08 205 871 and JP-A-08 217 796 and in the Japanese Patent 40 51 896.

Proteasome inhibitors are described for example in JP-A-05 000 968, WO 92/20 804, WO 94/17 816, WO 95/24 914, WO 95/25533, WO 96/13 266, WO 96/32 105 (lactacystin analogues) and US-A-55 80 854 (peptide aldehyde inhibitors).

Klafky et al. ((1995), Neuroscience Letters 201, 29-32) examine the effect of the proteasome inhibitor calpain inhibitor 1 on the secretion of β amyloid peptide which is formed by cleavage of the β amyloid precursor protein (APP) and which has been discussed to be a trigger of Alzheimer's disease. There are no proteasome structural data in this publication.

Fenteany et al. ((1995), Science, vol. 268, 726-731) describe the streptomyces metabolite lactacystin as a cell cycle inhibitor and inducer of neurite outgrowth of a mouse neuroblastoma cell line. The 20S proteasome was identified by means of tritium-labelled lactacystin as the specific cellular target of this inhibitor. A crystallizable proteasome preparation is not described.

WO 91/13904 describes the identification and characterization of a chymotrypsin-like protease which is present as a multicatalytic protease and its use for treating Alzheimer's disease. The use of substrates that are specific for chymotrypsin activity to test or screen

for inhibitors as described in this document only leads to the identification of inhibitors that are specific for a chymotrypsin-like activity.

Hence it is apparent that there is a great need for further information about proteasomes especially with regard to their exact structure in order to enable the preparation of new proteasome inhibitors in a rational manner. Thus the object of the invention was to provide a process which enables the crystallization of eukaryotic proteasome preparations so that the crystal structure can be used to simplify the development of new inhibitors.

The inventive object is achieved by a process for isolating a purified eukaryotic proteasome preparation comprising the steps:

- (a) production of a crude extract by lysing eukaryotic cells,
- (b) separation of insoluble components from the crude extract,
- (c) chromatographic separation into fractions by means of an ion exchange medium, e.g. Q-Sepharose,
- (d) testing the fractions obtained in step (c) and collecting the active fractions,
- (e) chromatographic separation over hydroxyapatite,
- (f) testing the fractions obtained in step (e) and collecting the active fractions,
- (g) concentrating the pooled fractions,
- (h) chromatographic separation over a gel filtration medium in a molecular weight range of 5 kD to 5 MD, e.g. Superose and
- (i) testing the fractions obtained in step (h) and collecting the active fractions.

Any eukaryotic cells can be used as a starting material for the process according to the invention e.g. animal cells, plant cells or fungal cells such as yeast cells. The use of yeast cells e.g. *Saccharomyces cerevisiae* is particularly preferred.

The fractions are usually tested during the purification process by determining the typical proteolytic activity for proteasomes. In this connection known chromogenic peptides can for example be used as substrates. The fractions are preferably tested by carrying out two parallel determinations of the proteolytic activity in each case, one of which is carried out in the absence and the other in the presence of a proteasome inhibitor e.g. lactacystin. This type of testing allows the fractions containing proteasomes to be unequivocally differentiated from other fractions with proteolytic activity.

The concentration process comprises three chromatographic separation steps (c), (e) and (h) of which at least one can be carried out in a FPLC system e.g. step (h).

A purified proteasome preparation is obtained by the process according to the invention which is present in an adequate amount and purity to enable a subsequent crystallization.

Hence a further subject matter of the present invention is a purified eukaryotic proteasome preparation which is obtainable by the process according to the invention. Yet a further subject matter of the present invention is a purified eukaryotic proteasome preparation in a

00364286-112026

crystallizable form. Yet a further subject matter of the present invention is a purified crystallized eukaryotic proteasome preparation.

The crystallized proteasome preparation can also contain a proteasome inhibitor. Examples of suitable known proteasome inhibitors are lactacystin or analogues thereof and tripeptide aldehydes such as calpain inhibitor.

The eukaryotic proteasome preparation according to the invention comprises a 20S proteasome i.e. a complex of 28 subunits each of which contains two molecules of seven different α type subunits and seven different β type subunits. In addition the complex can also contain metal ions e.g. magnesium, solvent molecules e.g. water and other polypeptide components.

The purified eukaryotic proteasome preparation according to the invention can be used to identify and isolate new proteasome inhibitors. Data from the crystal structure of crystallized eukaryotic proteasome preparations are used in particular for this. The identification and isolation of new proteasome inhibitors is preferably carried out using a computer-aided modelling program.

For example the inhibitor can be designed by visually inspecting graphic representations of the structure and in particular by

(a) determining the accessible volumes for ligands at active sites e.g. with the aid of the programs INSIGHT, SYBYL, QUANTA, FRODO, O etc.,

(b) determining ideal ligand properties with regard to

6620014-00275650

hydrophobicity or hydrogen bonds e.g. with the aid of the programs LUDI, GRID, etc. or/and

(c) determining the electronic properties of surfaces that are accessible to ligands at the active sites e.g. with the aid of the program GRASP etc.

Alternatively or additionally it is also possible to determine ligands by automated ligand fragment docking or adaptation procedures eg. with the aid of the programs DOCK, LUDI, LEAPFROG etc.

For this purpose it is particularly preferable to use the crystal data for the proteasome subunits of the β type in particular for the proteasome subunits β 5/PRE2, β 1/PRE3 or/and β 2/PUP1 or homologous subunits from other eukaryotic proteasomes and neighbouring subunits thereof e.g. β 4/C11 or/and β 7/PRE4.

In order to design inhibitors of the human proteasome it is possible to modify the inventive crystal structural data of the yeast proteasome by homology modelling using known amino acid sequences of the human proteasome. Such a homology modelling can be carried out by molecular graphic programs such as O, INSIGHT, FRODO, etc.. In particular the present invention encompasses a homology modelling of the homologous active sites of the active monomers in general and in particular for the purpose of inhibitor design. The homology of the amino acid sequences of the yeast proteasome and of the human proteasome in the relevant regions is shown in Figure 1.

In addition it is intended to illustrate the invention by the following examples and figures.

Figure 1

Sub
S1

shows the homology between the amino acid sequences from yeast and humans coding for the active subunits of the proteasome; the β_1 /PRE3, β_2 /PUP1, β_5 /PRE2 subfamilies are indicated by the yellow, green and blue colour respectively; the residues of the S1 pocket which influence the specificity changes of the PRE3 subfamily after substitution of the human subunit Y by LMP2 after cytokine induction are shown in brown;

Figure 2

shows the topology of the 28 subunits of the 20S proteasome drawn as spheres,

Figure 3

shows the C^α chain positions of the subunits β_7 /PRE4, β_6 /C5, β_1' /PRE3, β_2' /PUP1 and β_3' /PUP3 in which the β -cis and β -trans- β interactions by contacts of insertion segments are highlighted,

Figure 4a to b

show electron density maps (contoured starting at 1σ) in similar orientations around THR1 with two $F_o - F_c$ coefficients after double averaging; the red parts of the model were omitted via the phasing. β_5 /PRE2 with the covalently-bound lactacystin (LACT) and the water molecule NUK (a) and β_7 /PRE4 with a part of a propeptide (b),

Figure 5

shows a scheme of the proposed chemical steps of autolysis and substrate hydrolysis. Generation of a processing intermediate by hydrolysis on the acidic β ring surface (A).

Generation of the completely processed active subunit via an acyl enzyme (B) and its hydrolysis (C). Michaelis complex of a substrate polypeptide (D). Cleavage on the β ring surface and formation of the acyl enzyme (E) associated with peptide cleavage. Acyl enzyme hydrolysis and release of the octapeptide product (F).

Figure 6a to c show the binding of the calpain inhibitor and the S1 pockets, $\beta_1/\text{PRE3}$ is shown in grey with the P1 contacting residues shown in red; $\beta_2/\text{PUP1}$ is shown in green and the inhibitor is shown in blue (a); $\beta_2/\text{PUP1}$ (b), $\beta_5/\text{PRE2}$ (c) have an analogous colour scheme;

Figure 7 shows the lower half of the β - β chamber. The main chain with the red circles for the carbonyl oxygens is shown for the C-terminal sections of the helices H2 of the seven β type subunits which define the β ring area. The intermediary processed and the unprocessed propeptides of the subunits $\beta_6/\text{C5}$, $\beta_7/\text{PRE4}$, $\beta_3/\text{PUP1}$ and $\beta_4/\text{C11}$ (green) and the calpain inhibitor (yellow) bound to $\beta_1/\text{PRE3}$, $\beta_2/\text{PUP1}$ and $\beta_5/\text{PRE2}$ are shown. Two magnesium ions that are located near to the β ring area are shown as silver circles; and

Figure 8 shows a surface view of the proteasome molecule cut along the cylinder axis. Three of the six calpain inhibitor molecules bound to $\beta_1/\text{PRE3}$, $\beta_2/\text{PUP1}$ and $\beta_5/\text{PRE2}$ are shown in red as space

filling models. The sealed α openings at the two ends of the particle, a few narrow side windows and the sharply cut inner β ring surfaces can be seen.

Examples

Example 1 Protein preparation and characterization

Yeast cells of *Saccharomyces cerevisiae* (Hefe-Mayr, Munich, Germany) were washed twice with ice-cold water and suspended in buffer A (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM NaN_3) in a weight ratio of cells to buffer of 2:3. The cells were disintegrated for 10 min in a grinder (Biomatik, Germany) with glass balls (diameter: 0.5 mm; volume ratio of glass balls to cell suspension: 3:2). The disruption of the cells was monitored microscopically.

After filtration the crude extract was centrifuged for 10 min at 10,000 $\times g$ in a Sorvall RC 2B centrifuge. The supernatant was again centrifuged for 45 min at 134,000 $\times g$ in a Ti-55.2 rotor (Beckmann). The lipids from the upper most layer were carefully removed and the remaining yellow solutions were combined. The protein concentrations were about 50 mg/ml.

Immediately after centrifugation the extract was applied to a Q-Sepharose column (5 x 20 cm) which had been equilibrated with 280 mM NaCl in buffer A. The column was washed with 280 mM NaCl in buffer A, the proteins were eluted with a gradient of 280 to 800 mM NaCl. The flow rate was 120 ml/h and 12 ml fractions were collected. The proteasome was eluted at 400 - 450 mM

NaCl. Chymotrypsin-like (CL), peptidylglutamyl-peptide-hydrolase (PGPH) and trypsin-like (TL) enzymatic activities were measured in all fractions.

In order to obtain the 20S proteasome the CL activity was again measured in all active fractions in the presence of lactacystin and the fractions with a reduced activity were collected. The combined fractions were diluted three-fold with water and applied to a hydroxyapatite column (3 x 10 cm) which had been equilibrated with 60 mM potassium phosphate, pH 7.5. The column was washed with 60 mM potassium phosphate pH 7.5 and eluted with a gradient of 60-300 mM potassium phosphate. The flow rate was 60 ml/h. 12 ml fractions were collected. The CL, PGPH and TL activity was measured in all fractions and the active fractions were combined.

The combined fractions were concentrated 20-fold by ultrafiltration using a AMICON YM30 membrane and the concentrate was applied to a Superose 6 column (1 x 30 cm) equilibrated with buffer A. The elution was carried out with a flow rate of 18 ml/h in buffer A. The proteasome eluted after 37 min. In this manner it was possible to obtain 50 mg crystallizable protein from 500 g yeast cells.

All preparative steps with the exception of FPLC were carried out at 4°C. The chromogenic peptide substrates were dissolved in dimethyl sulphoxide at a concentration of 1 mM. The proteolytic activity towards these substrates was determined according to Achtstetter et al. (1994), J. Biol. Chem. 259, 13344-13348. The chromogenic peptide substrates were obtained from Bachem.

(Bubendorf, Switzerland). Q-Sepharose and hydroxyapatite were obtained from Sigma and BioRad. The FPLC device, the MonoQ and Superose 6 column were obtained from Pharmacia (Freiburg, Germany), all other chemicals were obtained in the highest possible purity from Merck (Darmstadt, Germany).

Example 2 Crystallization

The crystals were grown in hanging drops at 24°C. The protein concentration which was used for crystallization was 40 mg/ml in 10 mM Tris/HCl (pH 7.5) and 1 mM EDTA. The drops were composed of 4 μ l of the protein solution and 2 μ l of a reservoir solution which contained 40 mM magnesium acetate, 0.1 M morpholinoethane sulfonic acid (pH 6.5) and 12 % 2,4-methylpentanediol. The crystals containing the inhibitor lactacystin were produced by immersion in a 1 mM lactacystin solution for 6 h. The crystals containing the inhibitor acetyl-Leu-Leu-norleucine (calpain inhibitor I, Boehringer Mannheim) were produced by immersion in a 5 mM calpain solution for 6 h. The crystallographic data are shown in Table 1.

Example 3 Crystallography

The crystals were very-well ordered and exhibited only a slight anisotropy. Thus a resolution of 0.24 nm was possible. The acetyl-Leu-Leu-norleucinal-inhibited crystals were slightly less ordered.

The anisotropy of the diffraction was corrected using the found structure factor amplitudes with those that were calculated from a model with isotropic temperature factors using XPLOR (Bruenger, 1992). The data sets were

obtained with the BW6 beam line at the DESY Hamburg using a synchrotron radiation of $\lambda = 0.11$ nm. The crystals were immersed in an antifreeze buffer (30 % MPD; 28 mM magnesium acetate, 0.1 M morpholinoethane sulfonic acid, pH 6.9) and frozen in a stream of 90°K cold nitrogen gas. The diffraction data were collected with a 300 mm Mar research imaging plate at a distance of 275 mm (LACT) or 280 mm (CAL). The X-ray intensities were determined using the MOSFLM computer program version 5.3 and the data reduction was carried out with CCP4 (Leslie (1992), Acta Cryst. D50, 760-763; Joint CCP4 and ESF-EACMB, Newslett. Protein Crystallogr. (Daresburg Laboratory Warington UK 26M Collaborative Computational Project Number 4 (1994).

A rotation function calculated at 0.5 nm resolution showed two peaks that were related to the crystal symmetry which indicates the presence of local diadic molecule axes at ψ 86° ϕ 90° and ψ 94° ϕ 90°. Their correlation values were half the value of the crystallographic diad that would have been expected for an almost ideal molecular two-fold symmetry. The *T. acidophilum* model was used for the Patterson search calculation using AMoRe (Navaza (1994), Acta Cryst. A50, 157-163) at a resolution of 0.35 nm. This showed that if one takes into consideration the D7 symmetry of the experimental model, there is a single solution with a correlation value of 0.32 and an R factor of 56 % compared to the next-highest peak of 0.28 and 57 %.

The *T. acidophilum* model was reduced to polyalanine with only a few conserved residues which remained in the α type subunit. This model yielded an R factor of 57.7 % and was used to calculate a $2F_o - F_c$ map at 0.24 nm with X-PLOR (Bruenger (1992), X-PLOR version 3.1). A system

for X-ray crystallography and NMR was used. The electronic density was averaged in real space using MAIN (Turk (1992), Dissertation, Technical University Munich) and using the local double axis in the present model ($\psi=85.1$, $\phi=90.8$, $\kappa=180.1$), transformed back, and a new density was calculated with $2F_o - F_c$ coefficients. After 10 adjusting cycles the quality of the map was good ($R_{back}=27.3\%$). The individual subunits were identified on the basis of their characteristic insertions, deletions and amino acid sequences and were incorporated into the map on an ESV-30 graphic system work station (Evans & Sutherland, Salt Lake City, Utah) using FRODO (Jones (1978), *J. Appl. Cryst.* 11, 268-272). A crystallographic refinement was carried out with X-PLOR (Bruenger, 1992) with energetic and double non-crystallographic symmetry constraints using the parameters described by Engh and Huber (1991), *Acta Cryst. A* 47, 392 - 400. In addition a scattering component for the solvent was calculated in order to correct for an anisotropic crystal arrangement and was incorporated into the calculation of the model during the refinement.

The final model has taken into account the inhibitor molecules lactacystin and acetyl-Leu-Leu-norleucinal, 18 magnesium ions and 1,800 water molecules. The R values are satisfactory and the standard geometry of the bonds and angles are excellent. The local molecular diadic symmetry is well-conserved which is also shown by the very low value R_{back} 13 % in the final step of the analysis. The increase in the R value by 3 % for data with a resolution of 0.28 nm compared to 0.24 nm is a result of the anisotropic crystal arrangement which impairs the data quality and of the limited incorporation of ordered solvent molecules.

Example 4 Characterization of the structure

4.1 Structure of subunits

The 14 genes cloned from yeast which code for components of the 20S proteasome can be divided into seven α type and seven β type subunits.

The β type subunits are synthesized as precursors which are processed into the mature forms present in the assembled proteasome. The mature β type polypeptides $\beta 2/\text{PUP1}$, $\beta 5/\text{PRE2}$ and $\beta 1/\text{PRE3}$ are obtained from their proforms by cleavage between Gly-1 and Thr1 with release of the active site Thr1, whereas $\beta 7/\text{PRE4}$ is cleaved between Asn-9 and Thr-8 and $\beta 6/\text{C5}$ is cleaved between His-10 and Gln-9 and are present as stable processing intermediates. $\beta 4/\text{C11}$ and $\beta 3/\text{PUP3}$ are not processed and begin with Met(-1) or Met(-9) respectively. The subunits PUP1, PRE2 and PRE3 are referred to as completely processed, the subunits PRE4 and C5 as partially processed and the subunits C11 and PUP3 as unprocessed.

All 14 subunits are present in the crystalline molecular structure at well-defined positions. They are almost completely defined by the electron density apart from a few chain termini and long insertion segments.

The electron density for the main chains is defined as follows in the α type subunits: $\alpha 2/Y7$: Thr5-Leu236, $\alpha 3/Y13$: Gly4-Gly237, $\alpha 4/PRE6$: Tyr8-Gln244, $\alpha 5/PUP2$: Arg10- Glu243 (7 residues of the insertion are not defined - Gly12 to Arg 126), $\alpha 6/PRE5$: Phe4 - Ile233, $\alpha 7/C1$: Gly5 - Asn241, $\alpha 1/C7$: Gly6 - Asp240.

In the β type subunits the electron density is defined

as follows: $\beta_3/PUP3$: Ser-8-Asp 193, $\beta_6/C5$: Gln-9-Asp 193, $\beta_4/C11$: Met-1-Gln192, $\beta_7/PRE4$: Thr-8-Ile211, $\beta_2/PUP1$: Thr1-Cys221, $\beta_1/PRE3$: Thr1-Leu196, $\beta_5/PRE2$: Thr1-Gly211.

All seven α and β type polypeptides have a characteristic β sandwich structure. It is composed of two five-strand antiparallel β -folded sheet structures with the overlying helical layers composed of the helices H3, H4, H5 and the underlying helices H1 and H2. However, they differ in the bends which vary in length by one or two amino acid residues, in long insertions which connect secondary structural elements and in the N-terminal and in particular in the C-terminal regions.

In the α type subunits α 2/Y7 has a long insertion loop between the strands S9 and S10 which is composed of a short α helix and a β strand. α 1/C7 has an extension of the helix H3 by two bends as a result of the insertion at G180. The subunits α 1/C7, α 3/Y13, α 4/PRE6, α 5/PUP2 and α 7/C1 have longer C-terminal helices H5 which protrude into the solution from the particle surface. The highly charged, mostly acidic C-terminal segments are unstructured.

In the case of the β type subunits with long insertions, $\beta 7/PRE4$ has a sharp bend between the helices H1 and H2 and an additional α helix with 2 bends at residue 145. $\beta 6/C5$ has an insertion of 17 residues between H3 and H4 with a complex folding and a short helix. $\beta 2/PUP1$ has a very long C-terminal extension whose last 11 residues are very disordered. The subunits $\beta 3/PUP3$ and $\beta 6/C5$ have short C-termini so that the helices H5 do not exist and the strands S10 are extended to enlarge the β -folded

sheet. The helix H5 exists in β 4/C11 but it is two folds shorter than in the case of *T. acidophilum*.

Many of these subunit-specific folds, insertions and N- and C-termini are involved in the contacts between subunits as discussed in the following.

4.2 The (C7, Y7, Y13, PRE6, PUP2, PRE5, C1; PRE3, PUP1, PUP3, C11, PRE2, C5, PRE4)₂ complex

Each of the seven α type subunits has two neighbours within the heptameric ring which exhibit α -cis interactions and one or two neighbouring β type subunits in another ring with α -trans- β interactions. In addition to the β -cis and β -trans- α interactions, the central β type subunits have one or two neighbouring β type subunits in the other β ring with β -trans- β interactions.

The general architecture of the quarternary structure is the same in the proteasome of *T. acidophilum* and yeast (Fig. 2): The N-terminal loop segment, helix H0 (residues 20 to 30), loop L, the loop connecting H2 and S5 and the strand S7 mediate α -cis interactions. The β -cis contacts which appear to be less close include the loop L, the N-end of the helix H1, the strand S7 and the bend linking the strand S8 and the helix H3. These contacts are derived from the D7 symmetrical precursor and are also found in the *T. acidophilum* proteasome. Despite the conserved architecture these contacts are specific for the respective subunits due to their specific amino acid sequences.

Their are many additional contacts which are absent in

T. acidophilum and result from sequences and sequence insertions which are generally characteristic for yeast and eukaryotes. Within the α rings, close α -cis contacts are made by the intertwined N-termini of the subunits $\alpha 1/C7$, $\alpha 2/Y7$, $\alpha 3/Y13$ and $\alpha 7/C1$ in the centre of the heptameric ring. Tyr8 which is conserved in all subunits plays a central role. Within the β rings there is a very specific contact between $\beta 2/PUP1$ and $\beta 3/PUP3$ which is mediated by the long C-terminal arm of PUP1 which includes PUP3 and almost touches the next but one neighbour $\beta 4/C11$. β -trans- α contacts are made by the helix H1 loop helix H2 motifs which interact with the same motifs of two neighbouring α subunits. This basic contact motif was also seen in the *T. acidophilum* structure (see Figure 4a by Löwe et al. (1995), Science 268, 3479-3486) but the insertion at residue 66 of $\beta 7/PRE4$ favours its association with $\alpha 6/PRE5$ and $\alpha 7/C1$. In the same manner the long insertion in $\alpha 2/Y7$ at residue 210 between strands S9 and S10 binds to $\beta 2/PUP1$ and couples this pair. Specific β -trans- β interactions are formed by the C-terminal arm of $\beta 7/PRE4$ which is intercalated between $\beta 2'/PUP1$ and $\beta 1'/PRE3$. The C-terminal segment of $\beta 5/PRE2$ interacts with $\beta 3'/PUP3$ and $\beta 4'/C11$ in a similar manner (Figure 3). The long insertion of $\beta 6/C5$ at residue 145 contacts subunit $\beta 3'/PUP3$ and the C-terminal arm of $\beta 2'/PUP1$.

Very specific β -trans- β interactions are mediated by magnesium ions: magnesium Y8 bridges the main chain carboxylate of Asp193 from β 6/C5 with the loop 162 to 167 of β 2'/PUP1. In the same manner the magnesium Y9 bridges the subunits β 3/PUP3 via Asp193 with β 5'/PRE3. In addition these carboxylate groups are ligands for other magnesium ions which are located in the loops 165 of β 4/PUP3 (magnesium W6) and β 6/C5 (magnesium W4) and

can play a role in stabilizing the subunit structure. The aspartate residues are completely covered and their side chains participate in charge-charge interactions with Arg 19 of $\beta 2'$ /PUP1 and Arg 19 of $\beta 5'$ /PRE2 which further strengthens the β -trans- β contacts. The β type subunits $\beta 1$ /PRE3 and $\beta 4$ /C11 are located at the single molecule diads of the yeast proteasome and are very similar to the dominant β -trans- β contact the residues 133-137 of the helix H3 of *T. acidophilum*.

18 magnesium positions were identified in the proteasome molecule of which 12 are located on the inner walls of the β - β chamber and which prove the acidic nature of this compartment which is discussed in the following. It can be seen that the numerous specific interactions between the subunits determine their specific and unequivocal positions within the proteasome.

4.3 The N-terminal threonine position

A catalytic system with Thrl, Glu17 and Lys33 was defined in the *T. acidophilum* proteasome by structural and mutation investigations (Löwe et al. (1995), supra and Seemüller et al. (1995), *Science* 268, 579-582).

The residues Ser129, Ser169 and Asp166 are located close to Thrl which are required for the structural integrity of this position but could also be involved in the catalysis. It was shown by mutagenesis that Asp166 in the proteasome of *T. acidophilum* is essential (Seemüller et al. (1996), *Nature* 382, 468-470). These residues are invariant in the active subunits PUP1, PRE2 and PRE3.

In addition a completely bound solvent molecule NUK was

found in all three subunits close to Thr10^γ and N, Ser1290^γ and N and Gly47N as shown as an example for the subunit β5/PRE2 in the lactacystin complex (Figure 4). This was not recognized in a lower resolution in the model of *T. acidophilum*. Thr1N has hydrogen bridges to Ser1680 and O^γ and Ser1290^γ. Thr10^γ has a hydrogen bridge to Lys33^β. Asp17 has hydrogen bridges via O^{δ1} to Arg19N and Gly170N and via O^{δ2} to Thr/Ser2N and Lys33N^β. In a similar manner Lys33N^β has three hydrogen bridges to Asp17O^{δ2}, Arg190 and Thr10^γ.

The pattern of hydrogen bridges leads one to assume that Asp17 as well as Lys33 are charged. Thr1N can form a hydrogen bridge to Thr0^γ and is presumably neutral, a state which is favoured by a nearby positively charged lysine residue. Such a charge distribution would also be expected from the respective standard pKa values. Thr1N is therefore very probably the proton acceptor, if Thr10^γ is part of an electrophilic centre. This is confirmed by the structure of the lactacystin complex which has an ester between lactacystin and Thr1 as a result of a β lactone ring opening after a nucleophilic attack by Thr10^γ. Thr1N is at exactly the position to serve as a proton shuttle from Thr10^γ to lactacystin-06'. An analogous reaction sequence is proposed for the hydrolysis of the C-terminal fluorophores of fluorogenic substrates where the proton transfer takes place in the amide nitrogen of the leaving group. The generated acyl enzyme is deacylated by the water NUK as shown in the sections D-E of figure 5. Alternatively or concurrently NUK could directly attack the peptide bond thus circumventing the intermediate I.

4.4 Inhibitor binding

$\beta_3/PUP1$, $\beta_1/PRE3$ and $\beta_5/PRE2$ have the inhibitor acetyl-Leu-Leu-norleucinal bound covalently to Thr10 γ presumably as a hemiacetal. It assumes a β conformation and fills the gap between strands which contain the residues 20 and 21 and 47 (allocated to the loop L in figure 3 by Löwe et al., 1995, *supra*) to which it is bound via hydrogen bridges which generates an antiparallel β folded sheet structure. The norleucine side chain extends into a pocket (the S1 pocket) the side of which is open towards a tunnel which leads to the particle surface. The leucine side chain in P2 is not in contact with protein and the leucine side chain in P3 is in contact with the neighbouring β -subunits. The S1 specificity pocket is mainly formed by the residues 20, 31, 35, 49, 53 i.e. Ala20, Val31, Ile35, Met45, Ala49, Gln53(K) in $\beta_5/PRE2$ (fig. 6c), Thr20, Thr31, Thr35, Arg45, Ala49, Gln53 in $\beta_1/PRE3$ (fig. 6a), Ser20, Cys31, His35, Gly45, Ala49, Glu53 in $\beta_2/PUP1$ (fig. 6b). The residue 45 forms the bottom of the pocket and appears to largely determine its character. Neighbouring subunits in the β rings also contribute to the S1 pockets and modulate their character: $\beta_2/PUP1$ in the case of $\beta_1/PRE3$ with His114, His116, Ser118, Asp120; $\beta_3/PUP3$ in the case of $\beta_2/PUP1$ with the residues Asp114, Asp120 and Cys118 and $\beta_6/C5$ in the case of $\beta_5/PRE2$ with Ser118, Asp114, Glu120 and Glu122.

Lactacystin is covalently bound to β 5/PRE2. This is in agreement with the observed chemical modification of subunit X of the mammalian proteasome (Fenteany et al., 1995, Science 268, 726-730) the homologue of PRE2. Its dimethyl side chain at C10 extends into S1 like a valine or leucine side chain but less deeply than the norleucine side chain of calpain. Lactacystin forms several hydrogen

bridges with atoms of the protein main chain LactN-Gly470, Lact04'-Gly47N, Lact09'-Thr21N, Lact06'-Thr1N. Since these latter interactions can also occur in β 2/PUP1 and β 1/PRE3 which form no covalent complexes with lactacystin, the S1 side group which binds in the hydrophobic S1 pocket of β 5/PRE2 appears to direct the formation of a covalent bond and its stabilization. Thus this side group is an important starting point for the development of inhibitors.

4.5 Specificity

β 5/PRE2 has a methionine residue at position 45 which is in contact with the branched side chain of lactacystin in the complex. In the calpain-inhibitor complex the norleucine side chain of calpain pushes the methionine side chain by up to 0.27 nm towards Ile35 which rotates out of the way. This concerted movement makes the S1 pocket more spacious. This is compatible with the observation that lactacystin inhibits the chymotryptic activity towards chromogenic substrates. In a similar manner the chymotryptic activity is reduced in proteasomes with a β 5/PRE2 mutant which cannot be processed from their proform (Chen & Hochstrasser (1996), Cell 86 961-972) and by a mutation in β 5/PRE2 where a substitution of Ala49 by Val in the S1 pocket limits the size (Heinemeyer et al. (1993), J. Biol. Chem. 268, 5115-5120). β 1/PRE3 has an arginine residue in position 45 at the bottom of the S1 pocket which is well suited for glutamate P1 residues. It is most probably the subunit that is associated with the peptidylglutamyl-peptide hydrolysis activity (PGPH) of the proteasome. However, the norleucine side chain also occupies this basic pocket in the calpain inhibitor complex. A high additional density peak was observed which is associated with the

guanidinium side chain and can be interpreted as a chloride or carbonate ion which compensates a non-equalized positive charge. β_2 /PUP1 has a glycine as residue 45 and consequently a spacious S1 pocket, the bottom of which is bordered by His35 and Glu53.

We conclude that β_5 /PRE2 contains the chymotryptic as well as the tryptic activity whereas β_1 /PRE3 contains the PGPH activity but that both pockets are adaptable with regard to size (PRE2) and polarity (PRE3). β_2 /PUP1 is suitable for very large P1 residues with a basic character. Mutation analyses have shown that substitutions in β_4 /C11 and β_7 /PRE4 influence the chymotrypsin-like and the PGPH activity (Heinemeyer et al., (1993) *supra*; Hilt & Wolf (1996), *TIBS* 21, 96-102, Hilt et al. (1993), *J. Biol. Chem.* 268, 3479-3486). These subunits are inactive but are located in the vicinity of the subunits β_5 /PRE2 and β_1 /PRE3 from both rings (figure 4). The substitution of Ser136 by the voluminous Phe in β_4 /C11 interferes with the β -trans- β contact on helix H3 between β_4 /C11 and β_5 /PRE2 and can interfere with the neighbouring Thr1 position, as presumably also the deletion of 15 C-terminal residues of β_7 /PRE4, which form extensive contacts with β_1 /PRE3 (fig. 3).

4.6 Propeptides and processing

Five β type subunits are synthesized using propeptides of different lengths of up to 75 amino acids which are cleaved during maturation. β_2 /PUP1, β_5 /PRE2 and β_1 /PRE3 exhibit an autolysis between Gly-1 - Thr1. This is a process which requires the presence of Thr1, Gly-1 and Lys33. We had already suggested an autolysis within the subunit in which Thr10 γ as a nucleophile attacks the

preceding peptide bond (Schmidtke et al. (1996), EMBOJ. 15, 6887-6898).

According to the crystal structure the water NUK is allocated a central role. It is ideally positioned to act as a base for the removal of a proton of Thr10^γ and to promote nucleophilic addition to the carbonyl carbon of Gly-1. There is no information about the position and orientation of the Gly-1-Thr1 peptide group in the completely processed subunits but we can derive them from partially processed or unprocessed subunits β 3/PUP3, β 6/C5 and β 7/PRE4 which have similar orientations. In these subunits Gly-10 is directed towards the positively charged Lys33N^ε and of Gly47N which form an oxygen anion hole in analogy to serine proteases in order to distribute the negative charge that is formed when the tetrahedral adduct is formed. A rearrangement to form the ester can take place after proton transfer from the water NUK to Thr1N and cleavage of the peptide bond. The nearby residues Ser1290^γ and Ser1690^γ support this reaction. Both hydroxyl groups are bound to Asp166 which is invariant in the active subunits via hydrogen bridges. NUK probably also participates in the ester hydrolysis as an attacking nucleophile which is ultimately incorporated into the product (fig. 5, sections a to c). The Gly-1 residue appears to be essential since a side chain at position -1 would interfere with the protein backbone at position 168 and would force a configuration which is unsuitable for autolysis.

When Thr1 is released the subunits become active. If the catalytic site is not intact as in the subunits β 3/PUP3, β 6/C5 and β 4/C11 in which Thr1 is absent, in β 7/PRE4 in which Lys33 is replaced by Arg and in constructed variants of LMP2, the mammalian homologue of β 1/PRE3

(Schmidtke et al. (1996), *supra*) and of PRE2 (Chen & Hochstrasser (1996), *supra*) an autolysis at residue 1 does not occur. β 7/PRE4 has both essential residues Gly-1 and Thr1 but in a configuration which differs strongly from that in which the active subunits are found since the Thr1 side chain is pushed away by the larger Arg33 which replaces the lysine residue (figure 4b). The finding of defects in the catalytic activity and in the processing demonstrate the structural lability of the Thr1 site which can be impaired by mutations of neighbouring residues of the same or neighbouring subunits. On the other hand it is also possible that an inactive mutant can become active in the vicinity of active subunits which is in accord with observations that *T. acidophilum* species which have a defect in processing are processed when coexpressed with wild-type protein (Seemüller et al. (1996), *supra*).

The propeptides play an essential role in the assembly of eukaryotic proteasomes which may be due to direct or indirect effects by participation in interactions between subunits and/or by stabilizing the structure of subunits. The observed structures of the processing intermediates of β 7/PRE4(M) and β 6/C5 and of the unprocessed propeptide β 3/PUP3 indicate that both effects occur since the propeptides are firmly bound to the residue of the protein and interact with other subunits e.g. propeptide β 7/PRE4 with β 1/PRE3 and residues 92 and 115 and propeptide β 6/C5 with β 7/PRE4 at 91 and 116.

4.7 Entry into and exit from the proteasome particle

Sub D2
The hydrolytic activity of the proteasome is associated with Thr1 and the β ring surfaces in the interior of the

β cavity that defines the hydrolytic chamber. The substrate must penetrate into the particle and the product must be released. In the case of the proteasome from *T. acidophilum* two entry openings with a diameter of about 1.3 nm are open at the ends of the cylindrical particles which are bordered by a ring surface of bend-forming segments Tyr126-Gly-Gly-Val of seven identical α subunits. The N-terminal residues 1 to 12 are disordered in this protein.

In contrast the hydrolytic chamber of the 20S proteasome of yeast is almost inaccessible. The N-termini of the subunits $\alpha 1/C7$, $\alpha 2/Y7$, $\alpha 3/Y13$, $\alpha 6/PRE5$ and $\alpha 7/C1$ protrude into the opening and fill it completely with several layers of closely interwoven sidechains (figure 8). Hence there is no access to the interior of the particle from the cylinder ends without a considerable rearrangement. There are some narrow side windows in particular at the interface between the α and β rings which are more permeable than in the *T. acidophilum* proteasome since smaller side chains are present there. These openings are mainly located between the tooth-like helix H1-bend-helix H2 motifs of the α - β interface (see figure 4a by Löwe et al. (1995), *supra*) and lead to the N-terminal threonine residues of the active centre. They are covered by polar and charged amino acid side chains which can move in order to generate openings of about 1 nm diameter and possibly allow the passage of unfolded stretched polypeptide chains. The 19S particle which is responsible for the ATP dependency and ubiquitin dependency of proteolysis by the proteasome is attached to the particles to form the 26S proteasome. The association leads to a strong activation of peptide hydrolysis (Hoffman and Rechsteiner (1994), *J. Biol. Chem.* 269, 1690-1695). The proteasome regulator PA28 is

bound in a similar manner to α type subunits (Kania et al. (1996), Euro. J. Biochem. 236, 510-516). It accelerates peptide cleavage and improves the antigen processing. Both regulatory factors could open the entry openings in a controlled manner *in vivo*.

4.8 Generation of MHC class I peptides

The 20S proteasome generates peptide products with a narrow length distribution, mainly octapeptide or nonapeptides, a size range which is optimal for binding MHC class I molecules (York & Rock (1996), Annu. Rev. Immunol. 14, 369-396). In vitro experiments have shown that peptides generated from intact proteins by 20S proteasomes are presented by MHC class I molecules (Dick et al. (1994) Immunol. 152, 3884-3894; Niedermann et al. (1996), Proc. Natl. Acad. Sci. USA 93, 8572-8577). In an in vivo experiment it was shown that proteasome inhibitors inhibit the MHC class I presentation of protein antigens (Rock et al. (1994), Cell 78, 761-771) and that the number of the MHC class I molecules present on the cell surface is regulated by the inducible proteasome subunits β 5i/LMP7 and β 1i/LMP2 as shown in mice with site-directed deletions of the genes coding for these proteins (Fehling et al. (1994), Science 265, 1234-1237). LMP2 and LMP7 replace the constitutively expressed subunits after IFN- γ stimulation.

MHC class I peptides usually have basic or hydrophobic C-terminal residues (see the review article by Engelhard (1994), *Curr. Opin. Immunol.* 6, 13-23). The LMP2/7 substitution presumably changes the distribution of peptides so that a larger proportion of the peptides preferred by MHC class I molecules is generated. LMP2

replaces γ the human homologue of $\beta 1/\text{PRE3}$, LMP7 replaces χ the homologue of $\beta 5/\text{PRE2}$. All members of this subfamily exhibit a high degree of sequence identity but $\beta 1i/\text{LMP2}$ has two conspicuous differences compared to $\beta 1/\text{PRE3}$ in the S1 pocket: $\text{Thr31} \rightarrow \text{Phe}$ and $\text{Arg45} \rightarrow \text{Leu}$. The substitution of Arg by Leu makes the pocket unpolar and the substitution of Thr by Phe makes it narrower so that the PGPH activity should be reduced and the chymotryptic activity should be increased if $\beta 1i/\text{LMP2}$ replaces the mammalian homologue for $\beta 1/\text{PRE3}$. This is actually observed (Gaczynska et al. (1993), *Nature* 365, 264-267; Driscoll et al. (1993), *Nature* 365, 262-264) when LMPs are induced by treatment with $\text{IFN-}\gamma$. The opposite effect is found in cell lines which lack the LMP2 and LMP7 genes and in mutant mice with a disruption of the LMP2 gene (Van Kaer et al. (1994), *Immunity* 1, 533-541). Replacement of the mammalian homologues of $\beta 5/\text{PRE2}$ and $\beta 2/\text{PUP1}$ by LMP7 and MECL1 does not directly influence the S1 pockets and their effect cannot be due to the change in the specificity in P1 as is found for LMP2.

In the yeast 20S proteasome the subunits $\beta 7/\text{PRE4}$ and $\beta 6/\text{C5}$ are partially processed at residues -8 and -9. As a result octapeptide or nonapeptide products are formed which are not released from the enzyme. Both peptides have similar conformations with a thickening which subdivides two sections with extended conformations which is similar to the conformation of MHC class I-bound peptides. The similarity is quantified with rms deviations for all atoms of 0.23 nm and for the C^α atoms of 0.13 nm by comparing the propeptide of $\beta 6/\text{C5}$ with a viral peptide nonamer in a complex with its MHC class I receptor (Madden et al. (1992), *Nature* 321-325) and

allows the conclusion that preferred local conformations play a role in the generation (by the proteasome) and presentation (by MHC class I molecules) of immune-dominant peptide epitopes.

093285-120293